AD-A119 878 CALIFORNIA UNIV BERKELEY STATISTICAL LAB F/G 12/1
FUNDAMENTAL RESEARCH IN MATHEMATICAL STATISTICS AND PROBABILITY--ETC(U)
SEP 82 L L CAM, E L SCOTT DAAG29-79-C-0093
UNCLASSIFIED ARO-13676.12-MA NL

END
DATE FILMED
11 82
DTIC

# FUNDAMENTAL RESEARCH IN MATHEMATICAL STATISTICS AND PROBABILITY; Stochastic Processes

Lucien Le Cam
and
Elizabeth L. Scott

September 24, 1982

DEPARTMENT OF THE ARMY
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, North Carolina 27709

CONTRACT NUMBER
DAAG29-76-G-0167
DAAG29-79-C-0093

Statistical Laboratory
University of California
Berkeley, California 94720



Approved for Public Release;
Distribution Unlimited.

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. AD-A119 878 | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle) FUNDAMENTAL RESEARCH IN MATHEMATICAL STATISTICS AND PROBABILITY; Stochastic Processes | | 5. TYPE OF REPORT & PERIOD COVERED FINAL 1 April 1976 to 30 June 1982 |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s) Lucien Le Cam and Elizabeth L. Scott | | 8. CONTRACT OR GRANT NUMBER(s) DAAG29-79-C-0093 DAAG29-76-G-0167 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Statistical Laboratory University of California, Berkeley Berkeley, California 94720. | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709 | | 12. REPORT DATE 24 September, 1982 |
| | | 13. NUMBER OF PAGES 14 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report) Unclassified |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

MATH. STATISTICS, PROBABILITY--Construction of estimates, Empirical measures, Competing risks, Comparison of treatments, Identifiability of processes

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

BRIEF DESCRIPTION of research findings for the period 1-4-76 to 30-6-82: The results refer to a number of topics in the construction of tests or estimates with applications to selection procedures for studies, weather modification, tests of fit for logistic models competing risks, identifiability in Markov processes. A list of papers published and a list of Ph.D. granted are appended.

2. FOREWORD

This is a report of our activities for the period 1 April 1976 to 30 June, 1982. We regret to report that during that period Professor Neyman passed away (5 August, 1981). He had been active on ARO Research Grants since 1954. His death marks the end of a glorious era in Statistics. It also marks the end of the Berkeley Statistical Laboratory as we knew it. The Laboratory is presently under the able directorship of Leo Breiman whose range of interests differs materially from that of Jerzy Neyman. Professors E.L. Scott and L. Le Cam continued working on the grant until its expiration on 30 June, 1982. The summary of research findings given below for J. Neyman has been prepared by Lucien Le Cam.

3. TABLE OF CONTENTS

5. BODY OF REPORT

A. L. Le Cam

Most of the papers published during this period reflect ongoing studies in the domain of asymptotic methods in statistical decision theory. Although some may look theoretical, they were in fact directly inspired by questions from applications. The "Reduction theorem for certain sequential experiments" arose from the need to extend Neyman's technique of $C(\alpha)$ tests to sequential schemes. It says that if for non-random stopping rules one can approximate the family of distributions by exponential families, then one can restrict all considerations to stopping rules that depend only on the statistics occuring in the exponential approximations. The system behaves as if one used stopping rules on processes with independent increments. For these there is a large body of theory.

The paper, with R. Traxler, on "Mixtures of Poisson" arose from a claim by P.A.P. Moran that Neyman's treatment through $C(\alpha)$ tests does not apply to this case. We show that the technique can be made to cover very general cases, more realistic and more general than those covered in previous literature. We also detail what happens when the "usual regularity conditions" are not satisfied. Two papers deal with limit theorems for empirical measures. The first was a relatively crude extension of usual results to the case of independent but not identically distributed observations. It was needed to study the robustness of various procedures (see Paul Wang's thesis). The second is a better attempt inspired by work of R. Dudley and D. Pollard. It gives inequalities that can be used in a variety of problems, such as density estimation, construction of well

behaved estimates, etc.

The "Extension of a theorem of Chernoff and Lehmann" (with C. Mahan and A. Singh) came from the construction of test of goodness of fit for a logit model of effect of insulin treatment of prediabetic pregnant women. It gives extensions of the results of Djaparidze-Nikulin, D.S. Moore and Rao-Robson. It also gives a family of criteria intermediate between the ordinary Pearson $\chi^2$ and the Rao-Robson modification. These are intended for use in situations where the relevant matrices are near singular, a situation that occurs in particular in the logit case mentioned above.

There is a stray paper on "Tumor Growth". It came from an observation by Bartoszynski et al that the commonly used birth and death model does not fit the observable facts. We show that it cannot fit the observations even if one allows each patient to follow his/her own arbitrary random clock.

A number of results obtained under the grant have not yet made their way to the printing press. Some will be included in a monograph, "Asymptotic methods in statistical decision theory". It is almost ready, waiting for a variety of corrections. The publishing houses of Springer, Wiley and Wadsworth have expressed interest in the work.

One of the important chapters, Chapter 11, has previously been sent to ARO. It deals with the general case of approximability by finite dimensional Gaussian experiments. Previous versions, based on so called LAN conditions, could not deal with some very common problems in which the rates of convergence depend on the unknown parameters (as in explosive branching processes or in some regression problems).

There are also some other new features. One is about the treatment of dependent observations through a martingale approach. Billingsley had used martingales in his monograph on "Inference in Markov Processes". Our attack, suggested in "On local and global properties in the theory of asymp-

totic normality of experiments" (published in Stochastic Processes and Related Topics, Editor M.L. Puri, Acad. Press, 1975) has become standard. A definitive version is now being prepared by Priscilla Greenwood and A.N. Shiryayev.

Another new feature refers to a construction of estimates through tests between Hellinger balls. Using results of L. Birgé, we show that these estimates behave rather well. However at the present time we can can compute them effectively only in very special situations. Copies of the monograph will be made available as soon as corrections have been entered.

B. J. Neyman

Since the inception of the grant in 1976 Jerzy Neyman published 30 papers. The preparation of 8 of them was supported by the grant. The papers continue Neyman's study of a variety of natural phenomena together with the development of the statistical methodology needed to investigate them. Particular effects were made on the theory and use of $C(\alpha)$ tests (see Sankhyā paper #10 ). In a study of weather modification, the problems are to find out whether cloud seeding modifies 1) the probability of rain or 2) the average precipitation or 3) both. Tests of $C(\alpha)$ type for that purpose were developed by Neyman with the help of E.L. Scott, Barry James, Kang Ling James and Paul Wang. Still on weather modification, investigation of a number of experiments has shown that the effect of seeding can either increase or decrease rainfall and that the effects can be very pronounced as far away as 100 km from the intended target.

The development of statistical methodology to separate the occasions of increase from those of decrease is expected to be continued by E.L. Scott. At the time of his demise Neyman jointly with E.L. Scott had a first draft

of a monograph on the problem. The monograph is being continued by E.L. Scott.

Neyman was very interested in problems of pollution and health with special emphasis on carcinogenesis. The paper with P.S. Puri gives a model for the action of radiation on cells in culture. It yields the joint distribution of the number of surviving health and surviving transformed cells according to various characteristics of the incident radiation. An attempt is being made (by P.S. Puri, L. Le Cam and I. Janssen) to fit the model to actual data from the Lawrence Berkeley Laboratory.

For the action of various agents Neyman worked on the problem of detection of synergisms. The usual analysis of variance technique has the curious property that 1) the tests would accept the hypothesis of no interaction (= no synergisms) but 2) the interactions are there strongly enough to reverse the signs of the "main effects". A number of students in the department worked on this problem under the direction of J. Neyman and R. Bohrer.

Another problem that arises in many investigations of stochastic processes is that of lack of identifiability. For instance, in the case of Markov processes with a finite number of states when not all transitions are observable, there may be the possibility of constructing several different models, all leading to the same observable probabilities, but with radically different implications for intervention policies. The situation was studied by Neyman, with Peter Clifford (Oxford) and Brian Yandell.

Neyman also reported further work on different forms of the mechanism of clustering as it occurs in the distribution of insects in field, the development of populations, the spread of communicable diseases, the explosion of

galaxies and a number of natural phenomena. In the process of the above mentioned studies Neyman introduced the concept of "outlier prone" and "outlier resistant" families of distributions. A short description is given in paper (8).

C. E. L. Scott

The subjects we have been working on are all efforts to have more realistic studies of problems arising in many fields. In many examples of comparison of two treatments, of studies of survival, of reliability studies, and so forth, there are two possible effects that may be acting, one early and the other long-term. Weather modification treatments may change the probability that precipitation will occur at all, and also given that there is some precipitation the treatment may increase (or decrease or leave unchanged) the amount of precipitation per storm. Equipment may fail almost immediately due to faulty construction or it may gradually wear out. In most applications, both effects are of interest but it is the combination that provides the total value of the treatment. The two effects may act in the same direction or they may be acting in opposition. Many researchers ignore the first effect, others only apply one of the standard tests. Our asymptotic studies, accompanied by Monte Carlo simulations corresponding to twelve suggested criteria (16), show that the criteria designed to be locally optimum by Neyman and Scott, James and James, and Wang, are nearly equally good and quite reasonable when the two effects are in the same direction. But when they are in opposition, the power of every test goes down and the procedure which neglects the first effect and also the use of the standard tests give invalid tests with very low power. Work is in progress to develop new methods, such as two-stage tests that will detect

that two effects are acting in opposition and then estimate the effects.

A slightly different situation arises when the two effects are competing throughout time, which is the competing risks problem. For example, if there are two causes of failure, say A and B, then after the item fails from A, it cannot later fail from B and conversely. In many applications, we want to estimate the net failure rate of each cause. This is not difficult if A and B are independent, but it may be impossible if A and B are dependent (unidentifiable). However, if A depends on B but not conversely, estimation may be possible. We are progressing on determining the restrictions under which the net risks can be estimated and on providing efficient estimators.

We may try to obtain more precision by using multivariate observations but without assuming normality (for instance, one variable may be discrete). Under what conditions on the underlying distribution will it be more efficient to use the multivariate variables simultaneously, rather than the two variables individually (with Skurnick)? In a related class of problems, two types of observations may be contemplated, one of which is less expensive in time and money but also less precise. How can one use the cheap, fast data to design a more efficient experiment with the costly data? There are many applications including using mutagenicity tests to design carcinogenicity tests.

It is not easy to adapt exterior information to obtain more efficient estimates. We showed (13) that Rubin's empirical Bayes procedure will not provide more efficient selection methods but rather will tend to be harmful. Studies of selection of admission and the prediction of performance,

are practical problems on which we have been working intensively. A series of draft papers (29, 31 and others) has been prepared evaluating the efficiency of the standard selection procedures for students coming to a university, examining the differential biases in the standard methods from the viewpoint of sex, ethnicity, economic level, urban vs. rural, field of study. We then turned to searching for alternative selection procedures that will be both more reliable and fairer. We have found several possiblities and are testing these on other universities and for other years. One paper (23) will be published this year, and others are in preparation. Further studies investigate the year to year fluctuations and trends in the performance of students using Box-Jenkins analyses. Also under study are the effects of certain differences in the high school from which the student came.

There are occasions when the experimenter does not follow exactly the designed experiment as directed. A case study of this problem and the resulting interpretation in a lengthy and expensive weather modification experiment is described in (6).

PUBLISHED PAPERS

1) "A reduction theorem for certain sequential experiments", Statistical Decision Theory and Related Topics. Acad. Press, New York, 1977 (Le Cam).

2) "A statistician's view of weather modification technology", Proc. Nat. Acad. Sciences, Vol. 74, #11, 1977, pp. 4714-4721 (J. Neyman).

3) "On the asymptotic behavior of mixtures of Poisson distributions", Z. F. Wahrscheinlichkectstheorie u.v.G., Vol. 44, 1978, pp. 1-45 (Le Cam and R. Traxler).

4) "Maximum likelihood, an introduction", Lecture Notes #18, University of Maryland, 1979 (Le Cam).

5) "A reduction theorem for certain sequential experiments II", Ann. of Statist., Vol. 7, #4, 1979, pp. 847-859 (Le Cam).

6) "Comment on a paper by R. Braham", J. Amer. Statist. Assoc., Vol. 73, #365, 1979, pp. 70-77 (S. Dawkins and E.L. Scott).

7) "Comment on a paper by R. Braham", J. Amer. Statist. Assoc., Vol. 74, #365, 1979, pp. 90-94 (J. Neyman).

8) "Developments in probability and mathematical statistics generated by studies in meteorology and weather modification", Commun. Statist. Theor. Meth., Vol. A8, #11, 1979, pp. 1097-1110 (J. Neyman).

9) "Comments on the special issue of Communications in statistics (Vol. A8, #10) concerned with weather modiciation experiments", Commun. Statist. Theor. Meth., Vol. A9, #9, 1980, pp. 965-992 (J. Neyman).

10) "C($\alpha$) tests and their use", Sankhyā, Series A., Vol. 41, 1979, pp. 1-21 (J. Neyman).

11) "Some memorable incidents in probabilistic/statistical studies", Hoeffding Symposium, Acad. Press, New York, 1980, pp. 1-32 (J. Neyman).

12) "Comments on the discussion at the workshop on the statistical design and analysis of weather modification experiments", Statistical Analysis of Weather Modification Experiments, (E.J. Wegman and D. DePriest, eds.) Marcel Dekker, New York, 1980, pp. 131-137 (J. Neyman).

13) "Rubin's empirical Bayes computations are not useful for law school admissions", J. Amer. Statist. Assoc., Vol. 75, #372, 1980, pp. 821-823 (E.L. Scott).

14) "A hypothetical stochastic mechanism of radiation effects in single cells", Proc. Royal Society, London, B(213), 1981, pp. 139-160 (J. Neyman and P.S. Puri).

15) "Limit theorems for empirical measures and poissonization, Essays in Honor of C.R. Rao (G. Kallianpur, P.R. Krishnaiah and J.K. Gosh, eds.) North-Holland, 1982, pp. 455-463 (Le Cam).

16) "Comparison of two treatments when there may be an initial effect", Essays in Statistical Science (Festschrift for P.A.P. Moran), J. Applied Probability, Vol. 19A, 1982, pp. 253-264 (E.L. Scott).

17) "On the risk of Bayes estimates", Statistical decision and related topics III, Vol 2, Acad. Press, New York, 1982, pp. 121-137 (Le Cam).

PAPERS TO APPEAR

18) "A remark on empirical measures", Festschrift for E.L. Lehmann, Acad. Press., Nov. 82 (Le Cam).

19) "An extension of a theorem of H. Chernoff and E.L. Lehmann", Festschrift for H. Chernoff, Wadsworth, June 83 (Le Cam).

20) "On some stochastic models of tumor growth and metastasis", Probability Models and Cancer, North-Holland, December 82 (Le Cam).

21) "Stochastic theory of epidemics-continuing efforts to achieve realism", Contemporary Mathematics , Amer. Math. Soc., 1982 (J. Neyman).

22) "Asymptotic methods in statistical decision theory", Monograph, to be published by either Wiley or Springer, 1220 pages (Le Cam).

23) "Berkeley studies on alternative criteria for admission", Proc. of Symposium on Alternative Criteria for Admission to the University (ed. J. Sandoval), 1982 (E.L. Scott and H. Song).

24) "Neyman as applied statistician - some personal recollections", Amer. Statist. Assoc. Social Sciences Section, 1982 (E.L. Scott).

OTHER MANUSCRIPTS

25) "Public health policy and basic research" 1978 (J. Neyman).

26) "Clustering: reminiscences of some episodes in my research activity", Pfizer Colloquium, University of Connecticut, 1979 (J. Neyman).

27) "Problems in the stochastic modeling of skin cancer and its increases", 1979 (E.L. Scott).

28) "Risk analysis of methyl mercury poisoning", 1979 (E.L. Scott).

29) "Prediction of academic performance", 1979 (E.L. Scott).

30) "Assessing the goodness of fit of the logistic model", 1981 (Le Cam, C. Mahan and L. Kwei).

31) "Predictive study of student performance at the University of California. I. Descriptive statistics; II. Equations for predicting freshman GPA; III. Study of the residuals", 1982 (P.G. Neville, E.L. Scott, P.G. Wakim).

Ph.D. Theses Prepared with Partial Support of the Grant

1) Errol Caby, "Convergence of measures on uniform spaces", 1977.

2) Ryozo Miura, "Adaptive rank estimates for the one sample problem", 1977.

3) Amy Davis, "Robust measures of association", 1977.

4) J. Vadiveloo, "On the theory of modified randomization tests for non parametric hypotheses", 1977.

5) Photis Nanopoulos, "Measures on the integers, with applications", 1977.

6) Joel B. Brodsky, "On estimating a common mean", 1978.

7) Ki-Mok Kim, "Orientation-shift model on the sphere", 1978.

8) Albert Y.L. Lo, "Some contributions to Bayesian nonparametric statistical inference", 1978.

9) Paul Wang, "Asymptotic robust tests in the presence of nuisance parameters", 1979.

10) Bernard Davis, "Graphical techniques in reliability theory", 1979.

11) Seongill Park, "Tests for normality and symmetry", 1979.

12) Neng-Hsin Chen, "On the construction of a non parametric efficient location estimate", 1979.

13) Robert M. Holmes, "Contributions to the theory of parametric estimation in randomly censored data", 1981.

14) Hae-Hiang Song, "Interaction Studies in competing risk analysis", 1981.

15) Jane Ling Wang, "Asymptotic minimax estimates for distributions with increasing failure rate", 1982.

16) Long Kwei, "A chi-squared type goodness of fit test with random cells for conditional probabilities", 1982.

Ph.D. Theses Prepared with Partial Support of the Grant Whose Main Support was from Other Sources

17) P. Gorini, "Approximation of empirical Bayes by Bayes procedures", 1978.

18) J. Reeves, "A statistical analysis and projection of the effects of divorce on future U.S. kinship structure", 1982.

19) C.W. Wong, "Transformation of independent variables in regression models", expected December 1982.

PERSONNEL SUPPORTED UNDER THE GRANT

Faculty

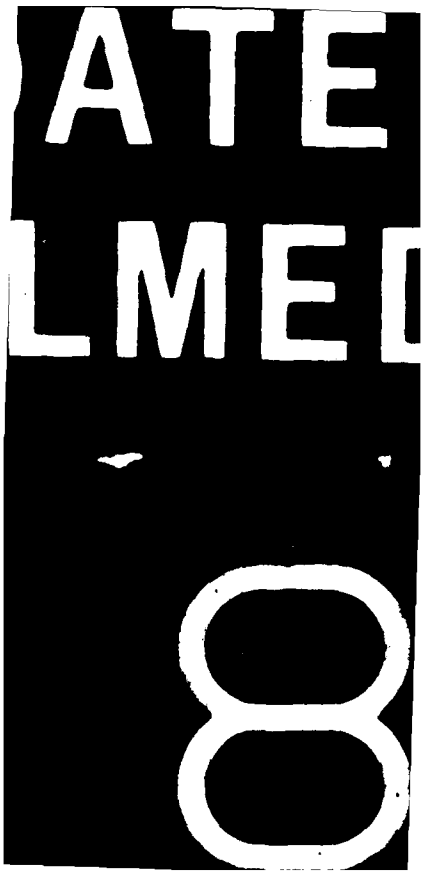
1) L. Le Cam, total approximately 6 months (Summer 1976 to 1981).

2) J. Neyman, total approximately 5 months (Summer 1976 to 1980).

3) E. L. Scott, total approximately 6 months (Summer 1976 to 1981).

Students

E. Caby, Ph.D.
J. Vadiveloo
K. Sharp, M.A. 1977
L. Kaiser
C. Wong, Ph.D.
S. Park, Ph.D.
P. Granger, M.A. 1977
E. Bailey
R. Holmes, Ph.D.
D. Shone (deceased)
N.H. Chen, Ph.D.
S. Vasudevan, Ph.D.
P. Gorini, Ph.D.
J. Morita
G. Ihaka, Ph.D. expected 1983
S. Hubert, Ph.D. expected 1983
J. Berger
B. Flannery
S.H. Lo, Ph.D.
M. Sobel
J. Reeves, Ph.D.
H. Song, Ph.D.
J.L. Wang, Ph.D.
L. Kwei, Ph.D.


ATE
LME
8